DOCUMENT RESUME

ED 193 318 TM 800 635

AUTHOR Bulcock, J. W.; And Others
TITLE Normalization Regression Estimation With Application to a Nonorthogonal, Nonrecursive Model of School Learning.
SPONS AGENCY Natural Sciences and Engineering Research Council, Ottawa (Ontario).; Social Sciences and Humanities Research Council of Canada, Ottawa (Ontario).
PUB DATE [80]
GRANT 86-786; A3960; A4353
NOTE 46p.; Contains occasional small print.

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Academic Achievement; Behavioral Objectives; *Foreign Countries; Grade 9; Grade 10; Learning Motivation; *Learning Theories; Least Squares Statistics; *Models; *Multiple Regression Analysis; Secondary Education; Student Characteristics
IDENTIFIERS *Bloom (Benjamin S); Canada; *Normalization Regression Estimation; Ridge Regression Analysis

ABSTRACT

Advantages of normalization regression estimation over ridge regression estimation are demonstrated by reference to Bloom's model of school learning. Theoretical concern centered on the structure of scholastic achievement at grade 10 in Canadian high schools. Data on 886 students were randomly sampled from the Carnegie Human Resources Data Bank. The 41 variable model formulated by Bloom demonstrated transmission of school performance and school-related attitudes from grades 9-10, as well as selected student background factors. Because school learning models call for reciprocal effects relationships at both entry behavior and learning outcomes stages, two stage least squares (2SLS) estimates were performed. Although incorporation of feedback linkages in structural equation models aggravated the multicollinearity problem to the extent that the 2SLS procedure was inadequate, normalization regression was able to cope satisfactorily with each of six harmful effets. The substantive results did not particularly support current theories of school learning, especially those concerned with the effects of socioeconomic status on schooling success, and the impact of sex differences on schooling outcomes. (Author/RL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED193318

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

NORMALIZATION REGRESSION ESTIMATION WITH APPLICATION TO A NONORTHOGONAL, NONRECURSIVE MODEL OF SCHOOL LEARNING *

BEST COPY AVAILABLE

J. W. Bulcock
Memorial University of Newfoundland

W. F. Lee
Memorial University of Newfoundland

W. S. Luk
Simon Fraser University

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
J. W. Bulcock

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

* Research supported by Grant No. 86-786 from the Social Sciences and Humanities Research Council; Grant No. A3960 and Grant No. A4353 from the Natural Sciences and Engineering Research Council

TM 800035

TITLE

NORMALIZATION REGRESSION ESTIMATION WITH APPLICATION TO A NONORTHOGONAL, NONRECURSIVE MODEL OF SCHOOL LEARNING

Abstract

Both discipline research and policy research are dependent on information of the kind: How responsive are behaviours Y_i to changes in X_i while taking other factors such as Z_i to Z_n into account? Research of this kind presents formidable problems of translation from theoretical (cause and effect) statements into research (probability) statements, and back again. As social science models increasingly call for the specification of feedback linkages the limitations of standard estimating methods in the presence of multicollinearity become apparent. It is argued, however, that the biased, reduced-variance approaches to the multicollinearity problem known as ridge regression analysis may prove inadequate where models are misspecified, as is commonly the case in the social sciences. Instead, a new approach known as normalization regression analysis is proposed. The advantages of the normalization regression method are demonstrated by reference to the estimation of a model of school learning formulated by University of Chicago scholar, Professor Benjamin S. Bloom.

1. Introduction

Discipline research is designed to advance knowledge in a scientific discipline. Policy research is designed to guide social action. Policy research can be either policy defining--that is, research designed to provide an information base for policy making in situations where the problem is recognized but where no policy action has been taken--or policy dictated--that is, research designed to evaluate the effectiveness of implemented policies. It is anticipated that in the post-industrial society research of both kinds will assume greater importance. (cf. Bell, 1973).

Both discipline and policy research are dependent on the accurate specification and estimation of models which can provide quantitative, explicit, and unambiguous answers to questions of the kind: How responsive are behaviours Y_1 to Y_n to changes in X_1 to X_n, other things such as Z_1 to Z_n being equal? Y_i behaviours may be any patterned social phenomena--inflation, unemployment, social mobility, poverty, school attainment, job status, the quality of school life, divorce, suicide, longevity and so on. In discipline research the X_i variables are usually the ones of theoretical interest, and the Z_i variables are the so-called control variables used to minimize spruriousness. In policy research the X_i variables are usually those which are amenable to change through policy action, whereas the Z_i variables are situational in that they are not readily amenable to policy control.

How the estimates of structural models are used depends on the type of research. The discipline-oriented analysts will use the results primarily for evaluating the authenticity of the theory, or the extent to which the specified relationships were shown to be false. The policy-

oriented analyst will use the results as a stepping stone to the conducting of structural analyses in order to answer policy questions of the kind:

(i) Which behaviours of interest are amenable to policy manipulation?

(ii) Will the policy manipulable variables make much difference over and above situational, or hard to change, variables?

(iii) What is the probability that the expected effect of changing policy variables will in fact occur?

(iv) How long will it take for the policy change to have the desired effect? and

(v) Will the proposed change have negative, neutral, or positive feedback consequences on other parts of interlocking behavioural systems?

The structural analysis techniques include comparative statics, multiplier analysis, and elasticities analysis.

As contemporary societies enter the post-industrial era they become more complex. Increasingly, simple, one-to-one cause-effect, solutions to complex problems are likely to fail; in fact, simplistic "solutions" may prove counterproductive. The increasing complexity of social systems is paralelled by research developments in the social sciences. The path of scientific development has proceeded from the formulation and estimation of simple models to more complex models. The tentative theory of experimental design developed by John Stewart Mill was formalized two generations later by Sir Ronald Fisher (1925, 1935); and the controlled experimental approach has given way in the past decade to structural equation model building.[1] (Cf. Duncan, 1975).

Thus, the earliest statistical models were designed to analyze the data of controlled experiments. These involved very few variables since the potential disturbances of contaminating influences were limited through radomization. But few social policy questions can be structured to fit the experimental mold. The testing of several treatments necessarily restricts the population to that included in the experimental design, and experimental subjects are seldom representative of the population of substantive research interest. Thus, the preferred realism of the natural setting is inevitably absent in controlled experiments.

As Kish (1975) points out "a research design always involves compromises of the desirable with the possible." And because the advantage of experimentation--control through randomization--is at the expense of the representativeness of samples which permits low level inferences to the defined population, compromise has taken the form of alternative methodologies. In the 1960's widespread use was made of ordinary least squares (OLS) estimating techniques as a single equation method for estimating recursive models of the interdependencies among set of variables. In the 1970's advances centered on the development of non-recursive single equation methods of which the two-stage least squares (2SLS) method has proven to be the most widely used of a variety of possible limited-information estimators (Intriligator, 1978, pp. 419-420). Flow diagrams, path models, and chain models became extensively used to depict model relationships and to facilitate model conceptualization.

Though it is claimed that these advances are being extensively used because most of their known statistical problems have been resolved (Land and Felson, 1978), they have given rise to a new generation of

problems (sui generis). For example, though the OLS regression method retains the Gauss-Markov property of minimum variance, in nonorthogonal models OLS estimates are biased; and, further, the small variance does not offset the large bias because of all estimating techniques OLS has the largest mean squared error (MSE). Thus, it is the least consistent in the sense that estimates do not automatically and indefinitely gain in accuracy if further observations are added.

On the other hand, 2SLS as a single equation approach performs better than OLS both in terms of bias and MSE. Further, it is less sensitive to specification error (that is, adding or deleting an explanatory variable) and measurement error, more consistent, and has the advantage of being able to estimate nonrecursive model parameters. Nevertheless, the 2SLS estimator is very sensitive to multicollinearity (King, 1974) though not necessarily more so than OLS estimation. Thus, when the Z_i (right hand side endogenous variables) in 2SLS equations are highly intercorrelated some of the resultant parameter estimates may be uninterpretable. What to do about this is the subject of the remainder of the paper.

For illustrative purposes we construct a model of school learning. We show that the nonrecursive features of the model aggravate the multicollinearity problem and that several 2SLS estimates defy interpretation. We discuss the possibility of a biased, reduced-variance solution of the ridge regression variety, but after examining several of the problematic properties of ridge regression reject the approach in favour of an alternative solution called normalization regression estimation. Because our focus is an applied one, all references to statistical theory

will be limited to footnotes. In this way it is hoped that the reader will neither lose sight of the substantive problem, nor be distracted by the statistical notation.

2. The School Learning Model

Bloom (1976, chap. 1) has formulated a micro-theory of school learning. Though the purpose of the theory is to explain the patterned variation of school learning, the motive underlying the theory is the specification of "an error-free system of education" (Bloom, 1976, p. 10). The school learning model, then, has both a discipline-orientation and a policy-orientation; that is, has both explanatory and policy-defining purposes.

We construct a model of school learning following Bloom's guidelines in order to illustrate some of the problems at the interface of theory and methodology that educational researchers are faced with. The model constructed for estimation purposes has three sets of variables; namely, (i) antecedent or background variables which constitute the exogenous variable set; (ii) a set of endogenous variables known as entry behaviours which are the student's lagged cognitive and affective learning characteristics; and (iii) another set of endogenous variables called the learning outcomes which are, again, comprised of both cognitive and affective measures.[2] Theoretical concern centers on the structure of scholastic achievement at the grade 10 (sixteen year-old) level in Canadian high schools.

The entry behaviours, as lagged endogenous variables, are treated as intervening between the influences of the background variables and the learning outcomes. The theory calls for an examination of the thesis that the cognitive and affective outcomes of schooling are mutually reinforcing--

a position congruent with that developed by social learning theorists. The argument is two pronged; namely, not only is behaviour today determined by an individual's behaviour yesterday, but also by an individual's perceptions of the consequences of that behaviour. In the context of the school learning model, the outcomes of grade 10 schooling will be influenced by the grade 9 entry behaviours and also by what the student anticipates the consequences of the grade ten outcomes will be. For example, students will be influenced in the present by their perceptions of future labour market opportunities (cf. Ogbu, 1978; especially chapters 5 and 6); and these influences will be reflected in their present motivations and expectations which in turn will affect their investments in cognitive competencies such as literacy and numeracy. Since success stimulates success cognitive performances will also be mutually reinforcing. Thus, the outcomes of grade 10 schooling constitute in theory a mutually reinforcing system of relationships.

Blalock (1979, p. 881) points out that "reality is sufficiently complex that we will need theories that contain upward of fifty variables if we wish to disentangle the effects of numerous exogenous and endogenous variables that interest us." School learning theory is no exception. In Table 1 we classify the 41 variables in the model to be estimated. Blalock (1979) also notes that realistic models must be nonrecursive; that is, must be able to cope with feedback properties. In Figure 1 we present a flow diagram of our general structural model and in Figure 2 the specified inner-relationships are shown which include reciprocal relationships between cognitive entry behaviours and the system of nonrecursive learning outcomes.[3] Thirdly, Blalock (1979, p. 881) draws

attention to the intractable problem of multicollinearity. Many of the variables in complex models will be highly intercorrelated, a problem which "will require a combination of large samples and good measurement". We cope with the measurement problem by following in the Noonan-Wold (1977, 1978) soft modelling tradition. The tradition makes extensive use of models involving latent variables which include within themselves much of the collinearity.[4] But we also show that this is not enough. The inner relationships between the latent constructs in the nonrecursive model are still highly collinear; hence, our consideration of biased, reduced-variance estimating methods.

TABLE 1, FIGURE 1, AND FIGURE 2 ABOUT HERE

3. Least Squares Estimation

Data

The data were obtained from the Carnegie Human Resources Data Bank. The Carnegie project was the first major study of Canadian education ever undertaken. It began as a five year panel study of all students (over 90,000) who began grade nine in Ontario high schools in 1959-60 and entered university in 1964. The data bank contains over 400 pieces of information on those students who stayed in school for the entire five year period--information on the student, his family, his teachers, his community and high school career. It is probably the largest panel study ever conducted.

The sample used for the analysis was selected as follows. A random ten percent subsample of about 9,000 was selected from the 90,000 grade nine population. Some 7,800 who spoke English at home were still in school in grade ten. Out of these cases there were no missing data on the 41 variables of interest for 886 respondents--443 males and 443

females. The coefficients for a list-wise matrix (N=886) were compared to those for a pair-wise matrix (N=7,809). There were few differences to the first decimal place. Thus, the case base of 886 was used in all statistical analyses.

Variables

See MacEachen (1960) and D'Oyley (1964) for the details concerning the measurement instruments used, and Weisbrod and Brehaut (1963) and Brehaut (1964) for information on the psychometric properties of the instruments. See Table 1 for a listing of the variables used in the school learning model. With the exception of the sex variable and two cognitive outcomes (English and mathematics achievement in grade 10) all the inner relations variables were latent constructs. Of these the socioeconomic status variable was treated as a productive or inward indicator model, and the remainder as reflective or outward indicator models.[5] The structures of the outer relations among the three sets of variables are shown in Figure 3.

FIGURE 3 ABOUT HERE

Ordinary Least Squares Results

If the assumption is made that the model is a recursive one, an ordinary least squares (OLS) regression analysis produces the Table 2 results. In Table 2 Dmax stands for the degree of multicollinearity in the model and can be interpreted as follows: zero for no collinearity and one for perfect collinearity.[6] It is concluded that in the entry behaviours segment of the model (the first four columns of Table 2) the degree of multicollinearity was negligible. At the learning outcomes stage of the recursive model (columns 5 through 8 of Table 2), however,

TABLE 2 ABOUT HERE

the Dmax values are higher by a factor of 17 than the Dmax values in the entry behaviours segment. The modest t-values at this stage, given the large case base, indicate that the standard errors of the estimates are unacceptably high. More important, however, is the fact that an OLS recursive model solution is unacceptable on the grounds that the model is functionally or mathematically misspecified. The hypothesized reciprocal relationships between literacy and numeracy at the grade nine level and the mutually reinforcing networth of learning outcomes relationships were not estimated. Feedback relationships call for 2SLS estimates which are presented next.

Two Stage Least Squares Results

Because school learning models call for reciprocal effects relationships at both entry behaviour and learning outcomes stages--see Figure 3--2SLS estimates are necessary. These are presented in Table 3. It is readily concluded from the Table 3 Dmax values that the functional respecification of the model through incorporation of feedback linkages has aggravated the multicollinearity problem. The respecification called for the estimation of the equations presented in Table 4. The reciprocal path results are for the most part uninterpretable. The t-values indicate that the standard errors are unacceptably high (cf. Table 4). Several negative paths are of dubious value (e.g., $p_{41,39}$; $p_{38,40}$; and $p_{39,41}$) because the estimated coefficients are larger than the underlying correlation coefficients between predictor and dependent variables; and the magnitude of several estimates appear inflated (e.g., $p_{41,40}$; $p_{38,39}$; $p_{40,39}$; and $p_{38,41}$). Finally, several residual paths have no substantive meaning (e.g., $p_{40,7}$; and $p_{3[illegible],6}$). The results illustrate all the symptoms of severe multicollinearity--a problem addressed next.

TABLES 3 AND 4 ABOUT HERE

4. Normalization Regression Estimation

The Multicollinearity Problem

Because nonorthogonal models in social science research are virtually unknown, the six harmful effects of multicollinearity are unavoidable. First, in the presence of multicollinearity, regression coefficients may have very large sampling variance.[7] Thus, the total variance; hence, the sampling variance of individual predictors, can be unacceptably large--even approaching infinity--when the data set approximates perfect collinearity.

Second, estimated coefficient vectors may be far too long in ill-conditioned data sets.[8] Thus, estimated coefficient vectors may be "off" in both magnitude and direction. This means that the estimated coefficients will be overly sensitive to the sampling error or to the functional mis-specification of the model due to the large variance; which, in turn, means that the inflated estimates will not accurately reflect the net relative importance of the variables under consideration. By the same token, it is possible for some non-significant coefficients to be interpreted as important because they are spuriously high. This is why estimated coefficients may be larger than the underlying correlation coefficient between the predictor and dependent variable; and why, occasionally, the OLS regression coefficients may be larger than one.

Third, when the data set is ill-conditioned OLS regression models do not provide accurate prediction; not even when the coefficient of determination (R^2) is high.[9] This accounts for the fact that high R^2 in collinear models does not produce accurate prediction; or why for that matter the prescience of the econometric forecaster is seldom equal to their important research tasks.

Fourth, OLS regression estimates are sensitive to sampling fluctuation when the data set is ill-conditioned. It has been noted (see Eq. 4, footnote 7) that for an ill-conditioned data set OLS estimates have large sampling variance due to their inverse relationship with the eigenvalues. Because the sampling variance approaches infinity when the data set approaches perfect collinearity the OLS estimates are sensitive to sampling fluctuation.

Fifth, estimated regression coefficients may be highly dependent; hence, error prone. When a data set has a high degree of multicollinearity, the estimates will have large variance and covariance; consequently they are likely to be highly correlated estimates. If one estimate is erroneous, because of their high dependency, the other will be erroneous. For example, when two predictors are highly positively correlated, the estimated coefficients will generally be highly negatively correlated; which means that if one coefficient is over-estimated, the other will be under-estimated. It is this high correlation between estimates, plus the large variance of the estimates which accounts for the excessively large sum of squares in OLS solutions and why the estimates may even have the wrong sign.

Sixth, OLS estimates are sensitive to model misspecification when the data set is ill-conditioned. Regression coefficients are not affected by model misspecification due to the erroneous inclusion or exclusion of variables when the data set is orthogonal. Under this condition, the data estimates are invariant of the model specification. But when the data set is ill-conditioned the betas are highly correlated and, hence, no longer invariant of the model specification; in which case the

regression coefficients of the included variables might change drastically as some relevant variables are added to, or irrelevant variables dropped from, the model.

Clearly, multicollinearity is an undesirable problem; one to be avoided if at all possible. But as Marquardt and Snee (1975) point out, in the social sciences it is impossible to impose control over all the environmental constraints accounting for intercorrelations between conceptually distinct predictor variables. Since all these harmful effects of multicollinearity were noted in the 2SLS solution to the Bloom model of school learning, and since as a consequence the model was uninterpretable, any procedure such as simple ridge regression (a biased reduced-variance approach) which promises a solution to these problem, deserves careful examination.

Simple Ridge Regression

Ridge regression analysis developed by Hoerl and Kennard (1970a, 1970b) is generally accepted as the most promising alternative to OLS regression for coping with collinearity. (cf. Dempster et al., 1977). In Hoerl and Kennard's simple ridge regression method the criterion for measuring the goodness of an estimator is the total mean square error (MSE)[10]. The MSE is the sum of the mean square error of the individual regression coefficients in an equation. It can be proved (e.g., Pindyck and Rubinfeld, 1976, p. 22) that the MSE can be decomposed into the sum of the total variance and the total squared bias of each regression coefficient.[11] Since, geometrically, MSE represents the expected squared distance between the estimated beta and the true beta, a "good" estimator would be one that minimized this distance--that is, minimized the MSE.

It follows that since the OLS beta estimate gives the minimum residual sum of squares; the ridge regression estimate based on the MSE criterion will give an inflated residual sum of squares.[12]

The ridge regression estimator, then, minimizes the squared length of the coefficient vector which is subjected to constant inflation in the residual sum of squares determined by the minimum MSE. In other words, the ridge estimator is a solution to minimizing the sum of squares when it is subjected to inflation--a constant determined by the maximum MSE.[13] The solution is obtained through Lagrangian optimization.[14] It is clear from Eqs. 12 through 14 (footnotes 12 through 14) that the minimum square length of the coefficient vector, or the minimum sum of squared regression coefficients determine the form of the ridge estimator (Eq. 12) and that the minimum MSE determines the value of the parameter k.

Most of the sixteen properties of ridge regression have been discussed by Hoerl and Kennard (1970a) and Marquardt (1970). See Lee (1979) for a summary. Three, however, are problematic. In the first place the total mean square error of the ridge regression estimator has two components. The first is the total variance of estimation which describes the random portion of the error, while the second is the square bias. But each is dependent on a true; hence, unknown, parameter, which means that the MSE cannot be obtained.[15]

Second, for any problem there exists a positive k_0 (called "optimum k") which gives a minimum MSE.[16] It is true even for an orthogonal system for which the degree of multicollinearity is zero. Third, for any problem the optimum k depends on the true regression coefficient vector and the

variance of the residual of the linear model.[17] But the multicollinearity problem arises from the interdependency among the explanatory variables and not from the dependency between the dependent variable y and the explanatory variables. Thus, the fact that k depends on the true regression coefficient vector and the variance of the residual which, in turn, are dependent on the y variable, means that the k value is necessarily stochastic when in fact and theory it should be nonstochastic.

There are four dilemmas of ridge regression stemming from these three properties of the procedure: namely, (i) the theoretical value of optimal k is stochastic when it should be nonstochastic if the task is to reduce the harmful effects of multicollinearity; (ii) the true optimum k depends on population parameters which are unknown--a fact which makes the problem unsolvable; (iii) even for an orthogonal system there is an optimum k at which MSE is munimum, and (iv) the bias of the ridge estimates can never be obtained. Thus, the optimum k cannot be calculated; consequently, the performance of ridge regression cannot be evaluated. So far more than fifteen methods have been described; for example, Hoerl and Kennard (1970a, 1976), Hoerl, Kennard and Baldwin (1975), Vinod (1976), Obenchain (1975), Hocking et al. (1976), McDonald and Galarneau (1975), Kasarda and Shih (1977), Hemmerle (1975), Hemmerle and Brantle (1978), Guilkey and Murphy (1975), Lawless and Wang (1976), Allen (1974). Unfortunately, despite their individual advantages none can give a better k or even a smaller MSE to that of OLS regression--a difficult that has marred the superiority of ridge regression over the OLS regression procedure.[18]

Normalization Regression

The total variance of a set of estimated regression coefficients can be shown to be a function of three factors:[19] (i) random error--that is, measurement error, stochastic error due to the incompleteness of the designed model, and some unknown inherent irreproducible fluctuation (Wannacott and Wannacott, 1969);[20] (ii) the sample size;[21] and (iii) the degree of multicollinearity, or the sum of the variance inflation factor (VIF) of each variable due to its interrelationship with the rest of the explanatory variables.[22]

From this relationship, it is deduced that for classical multiple regression the total variance of the estimated coefficient may be reduced by improving the precision of measurement, by improving model specification, by increasing the sample size, and most important by reducing the inflation of variance due to multicollinearity. The ridge regression procedure developed by Hoerl and Kennard (1970a) was originally designed to perform this last task--reduce the inflation of variance attributable to multicollinearity. But use of the MSE criterion for determining k is effectively a multifunctional procedure because in suppressing the VIF to less than unity it suppresses at the same time the variance attributable to other causes of variance--measurement error, model misspecification, and small sample size. It is this omnibus nature of the method which has forced ridge regression to be a stochastic procedure, and why even for an orthogonal system there is an optimum k which generates minimum MSE.

From the above argument, it follows that, if we want to limit the ridge procedure to the performance of the simple task of reducing the

inflation of variance due to multicollinearity, the lowest permissible total variance is that present in an orthogonal system.[23] In other words, if we wish to use a procedure which performs the task of reducing variance inflation due to multicollinearity, and only multicollinearity, we should normalize the average VIF such that it is equal to one. The value of k which satisfies this condition (see Eq. 21, footnote 23) is called the normalization k (or k_N).

If after using this procedure, the resultant variance is still too large, and if it is desirable to further reduce the variance, it should be accomplished through improvement in measurement error, model specification and sample size; but not by further suppression of the variance inflation factor; and especially if the intention is the interpretation of the phenomenon.[24]

Normalization Regression Results

The normalization regression results for the recursive model are presented in Table 5. In order to reduce the variance inflation due to multicollinearity the average VIF has been normalized such that it is approximately equal to one. The value of k_N which satisfies this condition is 0.020 at the entry behaviour stage, and 0.138 at the learning outcomes stage. There is a noticeable reduction in multicollinearity in the normalization regression model (see Table 5) compared to the OLS regression model in Table 2. The Dmax values were reduced by a factor of three and a half at the first stage of the model, and by a factor of four and a half at the second stage.

The normalization regression estimates as called for in the feedback effects model of school learning are shown in Table 6. These

TABLE 5 ABOUT HERE

results may be compared to those of the 2SLS model in Table 3 in order to ascertain the extent to which the dilemmas of multicollinearity have been coped with. The comparisons generate the following observations.

1. The introduction of modest bias designed to normalize the average VIF has significantly reduced the multicollinearity as indicated by the Dmax values.

2. The stronger t-values indicate significant reductions in the standard errors.

3. The magnitude of several estimates which appeared seriously inflated due to multicollinearity have been reduced.

4. The residual terms and coefficients are no longer uninterpretable.

5. In general, the 2SNR estimates were less sensitive to sampling fluctuation than the 2SLS estimates. See Table B, column 4 in footnote 24 for an example at the simple normalization regression stage of analysis.

6. Several nonsignificant negative paths in Table 3 which made little theoretical sense, turned out to have positive and statistically significant t-values in Table 6; hence, became interpretable in theoretical terms.

To claim that the normalization regression procedure solves the multicollinearity problem would be an overstatement. It reduces collinearity and copes with the problem in a manner which seems at least as satisfactory as other biased, variance-reduction techniques. (See footnote 24). Whether normalization regression has superior properties to ridge regression awaits the results of simulation experiments.

TABLE 6 ABOUT HERE

Attention is drawn, however, to three specific problems. First, compare 2SLS and 2SNR estimates for $p_{38,33}$; $p_{38,40}$; $p_{39,36}$; $p_{40,38}$; and $p_{41,39}$. The 2SLS beta estimates are too long; thus, they are "off" in both magnitude and direction, and are of the wrong sign. Second, the 2SLS paths $p_{38,32}$; $p_{38,39}$; $p_{39,40}$; and $p_{40,41}$, would have been interpreted as more powerful effects than in fact they were. Third, by the same token, some 2SLS estimates were spuriously low--e.g., $p_{38,40}$; $p_{40,34}$; $p_{40,38}$; and $p_{41,39}$--and would have been interpreted as non-significant when in fact they were significant.

Though concern in this paper is not with the substantive interpretation of the results, there are two findings of note. First, if the schools are class-biased it is not because social status membership directly depresses academic performance; but rather, because of its impact on student motivation and expectations which in turn affect academic performance. It seems that the relationship between social status and schooling is far more complex than current social theorists indicate. Second, the impact of social status on schooling is of modest proportions compared to the impact of sex differences. It appears that Ontario schools in the early 1960s may have been among the most sex biased in the industrial world. We do not know why. But the results show that the sex factor was a more powerful influence in determining school success than social status. Academically, girls had a decided advantage even though boys out performed girls in mathematics. And even though girls were significantly more highly motivated than boys, their expectations were lower. The winner was the highly motivated male student, who was a high performer in English language, and was from a high socioeconomic

background. The loser was the unmotivated female student, poor at English language, and from a low socioeconomic background.

5. Conclusion

In this paper we have shown:

(i) that the estimation of complex models may often call for solutions to formidable intellectual problems at the interface of theory and research;

(ii) that the formulation of dynamic, nonrecursive, structural models aggravates the problem of multicollinearity;

(iii) that though ridge regression methods cope with the estimation of ill-conditioned data more satisfactorily than OLS methods, they may still be unequal to the task because the mean square error criterion is stochastic, fails to satisfy the boundary condition, and is dependent on unknown and usually unknowable population parameters;

(iv) that, therefore, researchers in the model building tradition may wish to consider the application of a new criterion called the variance normalization criterion which was shown to cope with (but not solve) the multicollinearity problem as successfully as any other biased, reduced-variance technique; and

(v) that in some respects—notably in that (a) it suppresses the variance inflation of the estimates due only to multicollinearity, and not to other causes such as measurement error, model misspecification and small sample size, and (b) that it generates an unbiased OLS solution when the data set is orthogonal--it may prove a viable alternative to more traditional single equation estimating methods in so far as stochastic social science models are concerned.

The properties of normalization regression estimation were illustrated with reference to Bloom's (1976) model of school learning. The forty one variable model demonstrated the transmission of school performances and school-related attitudes from grades nine to ten while simulatenously taking into account selected student background factors. It was demonstrated that the incorporation of feedback linkages in structural equation models aggravated the multicollinearity problem to the extent that the 2SLS procedure was inadequate. Normalization regression--a biased, reduced-variance, collinearity suppression procedure--was shown to cope satisfactorily with each of the six harmful effects of multicollinearity. The substantive results did not particularly support current theories of school learning; especially those concerned with the effects of socio-economic status on schooling success, and the impact of sex differences on schooling outcomes.

NOTES

1. Though we use the phrase "structural equation model building," no generic name has emerged to define the family of methods designed to establish relationships between variables. In the psychological realm we can refer to psychometrics, but because in sociology the term has been preempted we cannot talk about sociometrics. Even terms such as biometrics, econometrics, psychometrics, and technometrics are ambiguous. For example, econometrics can refer either to the application of quantitative methods (or statistics) to economic problems, or to the methods of statistical analysis themselves. In the first case econometrics is a branch of economics; whereas in the second case econometrics is part of statistical theory. Despite content differences, the logic of the methods are the same in econometrics as in psychometrics, biometrics, or technometrics; all are primarily concerned with estimating the parameters of model structures. See Bielby and Hauser (1978) for a review of the structural equation tradition in sociology.

2. This formulation of the Bloom model may be seriously misspecified because a set of classroom-related variables has been omitted. School learning theorists such as Carroll (1963) and Bloom (1976) argue quite rightly that school achievement is a function of not only a student's prior learning behaviours and characteristics, but also a student's exposure to what is learned (the opportunity to learn, or task on time), and how what is to be learned is taught (the quality of instruction). These data were not gathered by the Carnegie project researchers. But if they had the present study would have been complicated even further. The units of measurement for all Table 1 variables were individuals, whereas opportunity to learn and quality of instruction variables are often classroom-level variables. Thus, had these variables been available to us we would have had to come to grips with the unresolved technical problem of linking both micro and macro-level theories in the same model.

3. In soft modelling the relations between the manifest or indicator variables (see Figure 3) are called outer relations, while the relations among the exogenous and endogenous variables are called inner relations. See the Noonan-Wold (1977) reference for a discussion of soft modelling.

4. Richard Noonan, a former student of Herman Wold's, has applied Wold's partial least squares techniques to ill-conditioned educational survey data. Latent variable modelling is a major breakthrough in social science research -- one that has been advanced by Karl Joreskog, another of Wold's former students.

5. Following Van de Geer (1971, formulae 13.13 and 13.15) factor score coefficients are used as the path coefficients in productive models and factor loadings as the path coefficients in reflective models.

6. Several indicators of the severity of multicollinearity have been used by different researchers (e.g., Price, 1977; Farrar and Glauber, 1976; White, 1978). Marquardt and Snee (1975) suggest that the maximum variance inflation factor, Vmax, is the best single measure. It is one for no collinearity and infinity for perfect multicollinearity. The measure used here, which gives 0 for no collinearity and 1 for perfect collinearity, is a transformation of Vmax given by,

$$Dmax = \frac{2}{\pi} \tan^{-1}(Vmax - 1) \qquad (1)$$

where the arc tangent is measured in radians. Because Dmax is the degree of multicollinearity of that variable affected the most, it is used to represent the multicollinearity in the equation matrix. Note that it is a relative measure which changes with the data manipulation technique. Thus, the Lmax for an OLS solution will be different (higher) than for a ridge regression solution. An absolute measure of multicollinearity "D_c" can be defined by using the Vmax of the principal component, $1/\lambda_{min}$; that is,

$$D_c = \frac{2}{\pi} \tan^{-1}(1/\lambda_{min} - 1) \qquad (2)$$

This measure represents the degree of multicollinearity in the data set and is invariant of the data handling technique. It is noted in passing that like the correlation coefficient D_{max} measures may take negative values. Positive Dmax signifies inflated variance (Vmax > 0) while negative Dmax signifies deflated or oversuppressed variance (Vmax < 0).

7. The total sampling variance for OLS estimates is given as follows (Hoerl and Kennard, 1970a):

$$\Sigma Var(\hat{\beta}) = \sigma^2 Tr(X'X)^{-1} \qquad (3)$$

$$= \sigma^2 \Sigma \frac{1}{\lambda_i} > \sigma^2/\lambda_{min} \qquad (4)$$

where Σ represents the summation from i=1 to i=p, the number of predictors, and λ_{min} is the smallest eigenvalue of the correlation matrix (X'X), which

approaches zero for data sets approximating perfect collinearity.

8. Geometrically, the total variance (see Eq. 3, footnote 7) can be regarded as the squared distance from the estimated coefficient vector, $\hat{\beta}$, to the true vector, β; and the greater the multicollinearity the longer the $\hat{\beta}$ to β distance. That is,

$$E\,(\hat{\beta}'\hat{\beta}) = \beta'\beta + \sigma^2 \Sigma \frac{1}{\lambda_i} > \beta'\beta + \sigma^2/\lambda_{min} \tag{5}$$

9. Though this follows from the above discussion regarding the instabilities and unreliabilities of regression coefficients due to multicollinearity the conclusion can be formulated explicitly.

Forecasting error variance, σ_f^2, of a model of the form $y = x\beta + \varepsilon$, can be written:

$$\sigma_f^2 = \sigma^2\{1 + \underline{x}'V\underline{x}\} \tag{6}$$

where $\underline{x}$ is a (p x 1) column vector of a single observation of the predictor variables, X's; V is the variance-covariance matrix of the estimated regression coefficients, which is $(X'X)^{-1}$ for OLS estimates, and σ^2 is the variance of the random error in y. Therefore, for OLS estimates the forecasting error can be expressed as:

$$\begin{aligned}\hat{\sigma}_f^2 &= \hat{\sigma}^2\{1 + \underline{x}'(X'X)^{-1}\underline{x}\} \\ &= \hat{\sigma}^2\{1 + \underline{x}^{*}{}'\Lambda^{-1}\underline{x}^{*}\} \\ &= \hat{\sigma}^2\{1 + \Sigma \frac{x_i^{*2}}{\lambda_i}\} \end{aligned} \tag{7}$$

$$\text{or, } \hat{\sigma}_f^2 = \frac{y'y(1 - R^2)}{n - p}\{1 + \Sigma \frac{x_i^{*2}}{\lambda_i}\} > \frac{y'y(1 - R^2)}{n - p}\frac{x_p^{*2}}{\lambda_{min}} \tag{8}$$

where $\underline{x}^{*} = P'\underline{x}$ and P is the orthogonal transformation matrix of (X'X), such that $P'(X'X)P = \Lambda$, the eigenvalue matrix of (X'X), and $P'P = PP' = I$. From Eq. (8) it is clear that the forecasting error variance can be very large for ill-conditioned data for which λ_{min} is very small, and that it is more sensitive to the smallest eigenvalue, λ_{min}, than to R^2.

10. The MSE function is defined by:

$$MSE = E\{(\hat{\beta}^{*} - \beta)'(\hat{\beta}^{*} - \beta)\} \tag{9}$$

Note that though two MSE criteria are used in ridge regression estimation, weighted and unweighted (Hocking et al., 1976), only the unweighted one is considered in this paper.

11. That is, $MSE = \Sigma Var\ (\hat{\beta}^*_i) + Bias^2(\hat{\beta}^*)$ (10)

12. The inflated residual sum of squares, ϕ, is written as:

$$\phi = \varepsilon'\varepsilon = (y - X\hat{\beta}^*)'(y - X\hat{\beta}^*)$$
$$= (y - X\hat{\beta})'(y - X\hat{\beta}) + (\hat{\beta}^* - \hat{\beta})'X'X(\hat{\beta}^* - \hat{\beta})$$
$$= \phi_{min} + \phi_0(\hat{\beta}^*) \quad (11)$$

where $\phi_0(\hat{\beta}^*) = (\hat{\beta}^* - \hat{\beta})'X'X(\hat{\beta}^* - \hat{\beta})$ (12)

is the inflation in the residual sum of squares. Geometrically, it gives the surface of a family of hyperellipsoids centered at $\hat{\beta}$, which is the OLS estimate of β.

13. That is, $\hat{\beta}^*$ is a solution to minimizing $\hat{\beta}^{*\prime}\hat{\beta}^*$ subjected to $(\hat{\beta}^* - \hat{\beta})'X'X(\hat{\beta}^* - \hat{\beta}) = \phi_0$ (a constant).

14. That is, by minimizing the Lagrangian function:

$$f(\hat{\beta}^*) = \hat{\beta}^{*\prime}\hat{\beta}^* + \frac{1}{k}\{(\hat{\beta}^* - \hat{\beta})'X'X(\hat{\beta}^* - \hat{\beta}) - \phi_0\} \quad (13)$$

where $\frac{1}{k}$ is the Lagrangian multiplier. At minimum $f(\hat{\beta}^*)$, we have:

$$\frac{\partial}{\partial\hat{\beta}^*} f(\hat{\beta}^*) + 2\hat{\beta}^* + \frac{1}{k}2X'X(\hat{\beta}^* - \hat{\beta}) = 0$$
$$(X'X + kI)\hat{\beta}^* = X'X\hat{\beta} = X'y$$

therefore,

$$\hat{\beta}^* = (X'X + kI)^{-1}X'y \quad (14)$$

where k is a parameter determined by Eq. (10), which in turn is determined by minimum MSE. (See Hoerl and Kennard, 1970a).

15. The ridge estimator $\hat{\beta}^*$ has a total MSE of:

$$MSE = \sigma^2\Sigma\frac{\lambda_i}{(\lambda_i + k)^2} + k^2\Sigma\frac{\alpha_i^2}{(\lambda_i + k)^2} \quad (15)$$

where $\alpha = P'\beta$ is the true regression coefficient vector in canonical form $y = X^*\alpha + \varepsilon$, and where $X^* = XP$.

Proof: The total mean square error

$$MSE = E(\hat{\beta}^* - \beta)^2 = E[\{\hat{\beta}^* - E(\hat{\beta}^*)\} + \{E(\hat{\beta}^*) - \beta\}]^2$$
$$= E\{\hat{\beta}^* - E(\hat{\beta}^*)\}^2 + \{E(\hat{\beta}^*) - \beta\}^2$$
$$= tr\ cov\ (\hat{\beta}^*) + Bias^2(\hat{\beta}^*)$$
$$= \sigma^2\ tr\ P\ Diag\{\lambda_i/(\lambda_i + k)^2\}P' + k^2\alpha'Diag\{1/(\lambda_i + k^2\}\alpha$$
$$= \sigma^2\Sigma\frac{\lambda_i}{(\lambda_i + k)^2} + k^2\Sigma\frac{\alpha_i^2}{(\ _i + k)^2}$$

16. Proof: (following Kasarda and Shi, 1977). Since

$$MSE = \Sigma Var\ (\hat{\beta}_i^*) + Bias^2(\hat{\beta}^*)$$

$$\frac{d}{dk}MSE = \frac{d}{dk}\Sigma Var(\hat{\beta}_i^*) + \frac{d}{dk}Bias^2(\hat{\beta}^*)$$

$$= -2\sigma^2\Sigma \frac{\lambda_i}{(\lambda_i + k)^3} + 2k\Sigma \frac{\lambda_i\alpha_i^2}{(\lambda_i + k)^3}$$

$$= -2\Sigma \cdot \frac{\lambda_i}{(\lambda_i + k)^3}\ \sigma^2 - k\alpha_i^2) \qquad (16)$$

Based on Rolle's Theorum (Widder, 1963), since

$$\lim_{k\to o} \frac{d}{dk} MSE = -2\sigma^2\Sigma \frac{1}{\lambda_i^2} < 0$$

the optimal k is a positive value. Further, since

$$\frac{d}{dk^2}MSE = 6\sigma^2\ \Sigma \frac{\lambda_i}{(\lambda_i + k)^4} + 2\Sigma \frac{\lambda_i\alpha_i^2}{(\lambda_i + k)^4}\ (\lambda_i - 2k) \qquad (17)$$

$$\lim_{k\to o} \frac{d^2}{dk^2} MSE = 6\sigma^2\Sigma \frac{1}{\lambda_i^3} + 2\Sigma \frac{\alpha_i^2}{\lambda_i^2} > 0 \qquad (18)$$

And according to the theory of minimum (Widder, 1963), this positive value leads to the conclusion that a minimum exists for MSE (Kasarda and Shih, 1977).

17. Proof: From Eq. 16 (footnote 16), at minimum MSE

$$\frac{d}{dk} MSE = -2\Sigma \frac{\lambda_i}{(\lambda_i + k)^3}\ (\sigma^2 - k\sigma_i^2) = 0 \qquad (19)$$

18. Some of the more positive properties of ridge regression have been overlooked here. Note, however, that: (i) ridge regression produces smaller variances of the regression coefficients than OLS regression (for proof, see Lee, 1979); (ii) ridge regression gives the minimum distance between $\hat{\beta}^*$ and the true vector β, which in this sense makes $\hat{\beta}^*$ a better estimator than $\hat{\beta}$, the OLS estimator (for proof, see Lee, 1979); (iii) though ridge regression produces a smaller multiple R^2 than OLS regression (for proof, see Lee, 1979), the ridge estimate gives a more accurate prediction equation than OLS regression if the bias introduced is not too large (for proof, see Lee, 1979); and (iv) from (i) we can imply that ridge regression is less sensitive to sampling fluctuation.

19. The total variance may be expressed as:

$$\Sigma \mathrm{Var}(\hat{\beta}^*) = \hat{\sigma}^2 \mathrm{tr}(X'X)^{-1}$$

$$= \frac{\varepsilon'\varepsilon}{n-p} \mathrm{tr}(X'X)^{-1} \qquad (20)$$

where ε is the random error vector of y in the standardized regression model, $y = X\beta + \varepsilon$, n is the sample size, p is the number of explanatory variables, and $\mathrm{tr}(X'X)^{-1}$ is the trace of the inverse of the correlation matrix (X'X).

20. The two errors; namely, measurement error ε_m, and stochastic error ε_s are assumed to be uncorrelated and to have normal distribution with zero mean. Thus, the sum of the squares of random error $\varepsilon'\varepsilon$ can be expressed as $\varepsilon'\varepsilon = \varepsilon_m'\varepsilon_m + \varepsilon_s'\varepsilon_s$. Obviously, random error can be reduced though not eliminated by improving the precision of measurement and by improving the designed model.

21. More specifically, sample size n, is the degree of freedom of the sum of the squares error (n - p). Thus, total variance can be reduced by increasing the sample size.

22. The effect of the multicollinearity enters through $\mathrm{tr}(X'X)^{-1}$; that is, the trace of $(X'X)^{-1}$.

23. The lowest possible variance is $p\hat{\sigma}^2$, which is the value of the best condition; that is,

$$\Sigma \mathrm{Var}(\hat{\beta}^*) = p\hat{\sigma}^2,$$

or, $$\hat{\sigma}^2 \Sigma \frac{\lambda_i}{(\lambda_i + k)^2} = p\hat{\sigma}^2$$

and therefore,

$$\frac{1}{p} \Sigma \frac{\lambda_i}{(\lambda_i + k_N)^2} = 1 \qquad (21)$$

where k_N is the normalization k.

24. Though it has been shown elsewhere (Lee, 1979; Lee and Luk, 1979) that the normalization procedure is more consistent in terms of performance than other ridge regression methods, the superiority properties of the normalization regression method may be demonstrated by reference to the Bloom model of school learning. Consider the recursive OLS prediction equation for grade 10 achievement in English language $\{X_{38} = f(X_{31} \text{ thru } X_{37})\}$.

The path coefficients for this equation for eight different estimating methods are presented in Table A and the characteristics and performance indices of each method are summarized in Table B.

Tables A and B about here

The comparisons between the path coefficients produced by OLS and the ridge regression methods are not especially interesting. With the sole exception of Vinod's method the substantive interpretation of the parameters of the recursive English language achievement model would probably be unaffected by the differences between the estimates. It seems that a 0.7 degree of multicollinearity need not prove problematic.

From Table B it is concluded that the normalization method: (i) was the only method to reduce the minimum VIF to a value close to unity (column 2); (ii) was the method which generated the maximum reduction of the relative degree of multicollinearity (column 3) without introducing too large a bias (column 1); (iii) was the one which produced the largest amount of reduction in the total variance of the estimates (column 5) without deflating or over-suppressing the variance (column 3); (iv) produced more stable estimates than OLS regression on the grounds that an E.S.Ratio of .405 shows that OLS is two and a half times more sensitive to sampling error than normalization regression analysis (column 4); (v) produced low inflation in the residual sum of squares; and (vi) was significantly different from OLS estimation as shown by the α-acceptability level.

Note (Table B, column 8) that ridge regression, including normalization regression, produces a smaller multiple R square than that of OLS regression. For proof see Lee (1979, p. 30). Nevertheless, the ridge estimate always produces a more accurate prediction equation than OLS regression if the bias introduced is not too large. Thus, the conventional wisdom that a reduction in R square indicates a poorer fit is definitely not correct. On this point see also Barrett (1974), and Pindyck and Rubinfeld (1976, p. 61). For proof, see Lee (1979, p. 31).

References

Allen, D.M. The relationship between variable selection and data augmentation and method for prediction. Technometrics, 1974, 16, 125-127.

Barrett, J.P. The coefficient of determination--some limitations. The American Statistician, 1974, 28, 19-20.

Bell, D. The coming of post-industrial society: A venture in social forecasting. New York: Basic Books, 1973.

Bielby, W.T. & Hauser, R.M. Structural equation models. In A. Inkeles, J. Coleman, and N. Smelser (Eds.), Annual Review of Sociology, Volume 3. Palo Alto, CA: Annual Reviews Inc., 1978.

Blalock, H.M.Jr. Measurement and conceptualization problems: The major obstacle to integrating theory and research. American Sociological Review, 44, 1979, 881-894.

Bloom, B.S. Human characteristics and school learning. New York: McGraw-Hill, 1976.

Brehaut, W. A preliminary look at the Carnegie students in grade 12 in Ontario schools: Analysis of the replies from the student and staff questionnaires. Carnegie Study Bulletin No. 8, Department of Educational Research, University of Toronto, 1964.

Carroll, J.B. A model of school learning. Teachers College Record, 64, 1963, 723-732.

Dempster, A.P., Schatzoff, M. & Wermuth, N. A simulation study of alternatives to ordinary least squares. Journal of the American Statistical Association, 72, 1977, 77-91.

D'Oyley, V.R. The Canadian tests, grades 8, 9, and 10. Carnegie Study Bulletin No. 5, Department of Educational Research, University of Toronto, 1964.

Duncan. O.D. Introduction to structural equation models. New York: Academic Press, 1975.

Farrar, D.E. & Glauber, R.R. Multicollinearity in regression analysis: The problem revisited. Review of Economics and Statistics, 1967, 92-107.

Fisher, R.A. Statistical methods for research workers. Edinburgh: Oliver and Boyd, 1925.

Fisher, R.A. The design of experiments. Edinburgh: Oliver and Boyd, 1935.

Guilkey, D.K. & Murphy, J.L. Directed ridge regression techniques in case of multicollinearity. Journal of the American Statistical Association, 70, 1975, 769-775.

Hemmerle, W.J. An explicit solution for generalized ridge regression. Technometrics, 17, 1975, 309-314.

Hemmerle, W.J. & Brantle, T.F. Explicit and constrained generalized ridge estimation. Technometrics, 20, 1978, 109-120.

Hocking, R.R., Speed, F.M. & Lynn, M.J. A class of biased estimators in linear regression. Technometrics, 12, 1976, 55-67.

Hoerl, A.E. & Kennard, R.W. Ridge regression: Biased estimation for nonorthogonal problems. Technometrics, 12, 1970a, 55-67.

Hoerl, A.E. & Kennard, R.W. Ridge regression: Application to non-orthogonal problems. Technometrics, 12, 1970b, 69-82.

Hoerl, A.E. & Kennard, R.W. Ridge regression: Iterative estimation of the biasing parameter. Communications in Statistics - Theoretical Methods, A5(1), 1976, 77-88.

Hoerl. A.E., Kennard, R.W. & Baldwin, K.F. Ridge regression: Some simulation. Communications in Statistics, 4(2), 1975, 105-123.

Intriligator, M.D. Econometric models, techniques, and applications. Englewood Cliffs, N.J.: Prentice-Hall, 1978.

Kasarda, J.D. & Shih, W.F. Optimal bias in ridge regression approaches to multicollinearity. Sociological Methods and Research, 5, 1977, 461-470.

King, D. Feedback and reciprocal causation: A Monte Carlo study of the relative small sample properties of small sample estimators. Unpublished doctoral dissertation, University of North Carolina, 1974.

Kish, L. Representation, randomization, and control. In H.M. Blalock, Jr. et al. Mathematical models and international perspectives on mathematical sociology. New York: Academic Press, 1975.

Land, K.C. & Felson, M. Sensitivity analysis of arbitrarily identified simultaneous-equation models. Sociological Methods and Research, 6, 1978, 283-307.

Lawless, J.F. & Wang, P. A simulation study of ridge and other regression estimators. Communications in Statistics--Theoretical Methods, A5(4), 1976, 307-323.

Lee, W.F. Model estimation using ridge regression with the variance normalization criterion. (Interim Report No. 2, Grant No. S76-0645, The Social Sciences and Humanities Research Council of Canada). St. John's, Newfoundland: Memorial University of Newfoundland, 1979.

Lee, W.F. & Luk, W.S. Simple ridge regression with the variance normalization criterion. In J.W. Bulcock (Ed.), Simple ridge regression with the normalization criterion. (Interim Report No. 1, Grant No. S76-0645, The Social Sciences and Humanities Research Council of Canada). St. John's, Newfoundland, Memorial University of Newfoundland, 1979.

McCabe, G.P. Jr. Evaluation of regression coefficient estimates using α-acceptability. Technometrics, 20, 1978, 131-139.

McDonald, G.C. & Galarneau, D.I. A Monte Carlo evaluation of some ridge-type estimators. Journal of the American Statistical Association, 70, 1975, 407-416.

MacEachern, D.G. Twenty questions. Carnegie Bulletin No. 1, Department of Educational Research, University of Toronto, 1960.

Marquardt, D.W. Generalized inverse, ridge regression, biased linear estimation, and nonlinear estimation. Technometrics, 12, 1970, 591-612.

Marquardt, D.W. & Snee, R.D. Ridge regression in practice. The American Statistician, 29, 1975, 3-20.

Noonan, R. & Wold, H. NIPALS path modelling with latent variables: Analyzing school survey data using nonlinear, iterative, partial least squares. Scandinavian Journal of Educational Research, 21, 1977, 33-61.

Noonan, R. & Wold, H. PLS path modelling with latent variables: Analyzing school survey data using partial least squares, part II. Working paper No. 1, Institute of International Education, University of Stockholm, Sweden, 1978.

Obenchain, R.L. Ridge analysis following a preliminary test of the shrunken hypothesis. Technometrics, 17, 1975, 431-445.

Ogbu, J.U. Minority education and caste: The American system in cross-cultural perspective. New York: Academic Press, 1978.

Pindyck, R.S. & Rubinfeld, D.L. Econometric methods and economic forecasts. New York: McGraw-Hill, 1976.

Price, B. Ridge regression: Application to non-experimental data. Psychological Bulletin, 84, 1977, 759-766.

Van de Greer, J. Introduction to multivariate analysis for the social sciences. San Francisco: W.H. Freeman, 1971.

Vinod, H. D. Application of new ridge regression methods to a study of Bell system scale economics. Journal of the American Statistical Association, 71, 1976, 835-841.

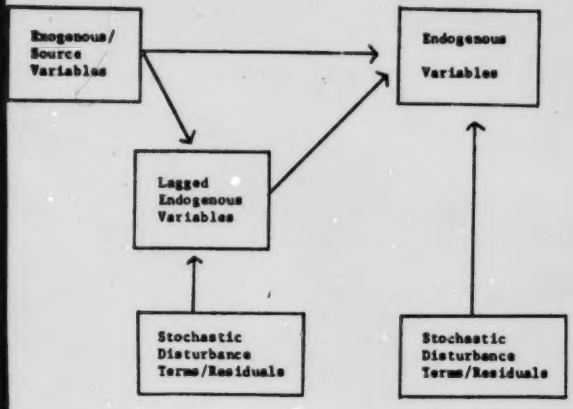
Wannacott, R.J. & Wannacott, J.H. Econometrics. New York, Wiley, 1969.

Weisbrod, V.M. & Brehaut, W. School and general health characteristics of grade 9 students in Ontario. Carnegie Bulletin No. 2, Department of Educational Research, University of Toronto, 1963.

White, V.J. A general computer program for econometric methods: SHAZAM. Econometrica, 46, 1978, 239-240.

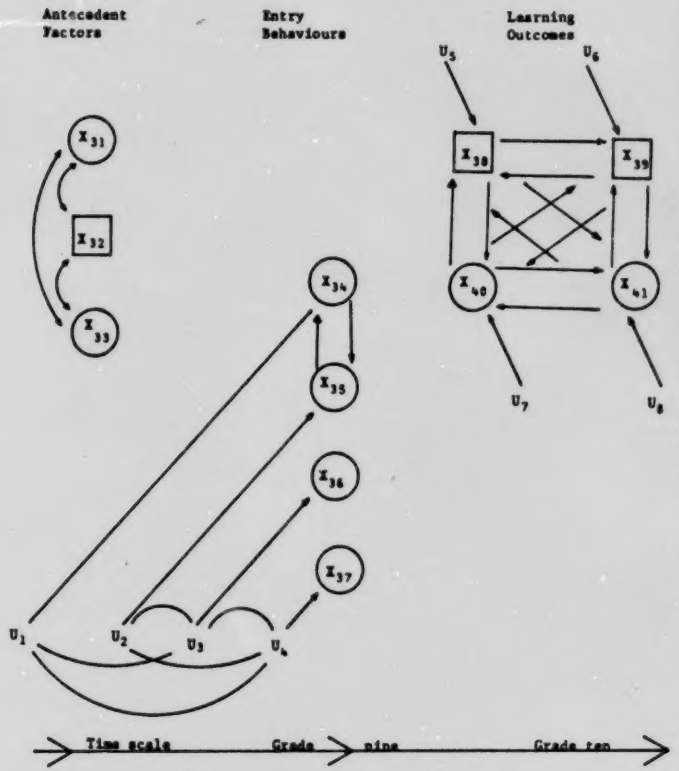
Widder, D.V. Advanced calculus. Englewood Cliffs, N.J.: Prentice-Hall, 1963.

FIGURE 1. A FLOW DIAGRAM OF THE GENERAL STRUCTURAL MODEL[a]



a. A description of the structural form and reduced form of the general structural model is given by Intriligator (1978, pp. 39-45).

FIGURE 2. THE STRUCTURE OF INNER RELATIONS FOR THE BLOOM MODEL OF SCHOOL LEARNING[a]



a. Recursive path arrows were omitted in the interests of clarity. Variables are assumed to be caused by the variables to the left of them. Latent variables are circled; observed variables are in boxes. See Table 1 for variable identification.

FIGURE 3

THE STRUCTURE OF OUTER RELATIONSHIPS[a]

Figure 3a. Outer Relations among Antecedent Variables

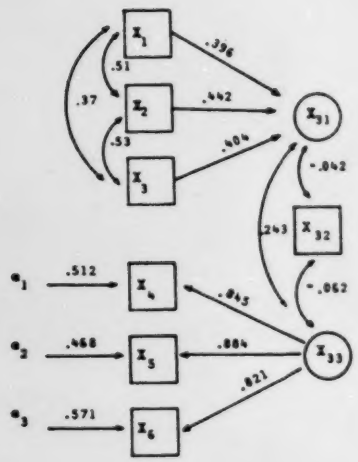


FIGURE 3b. Outer Relations Among Entry Behaviours

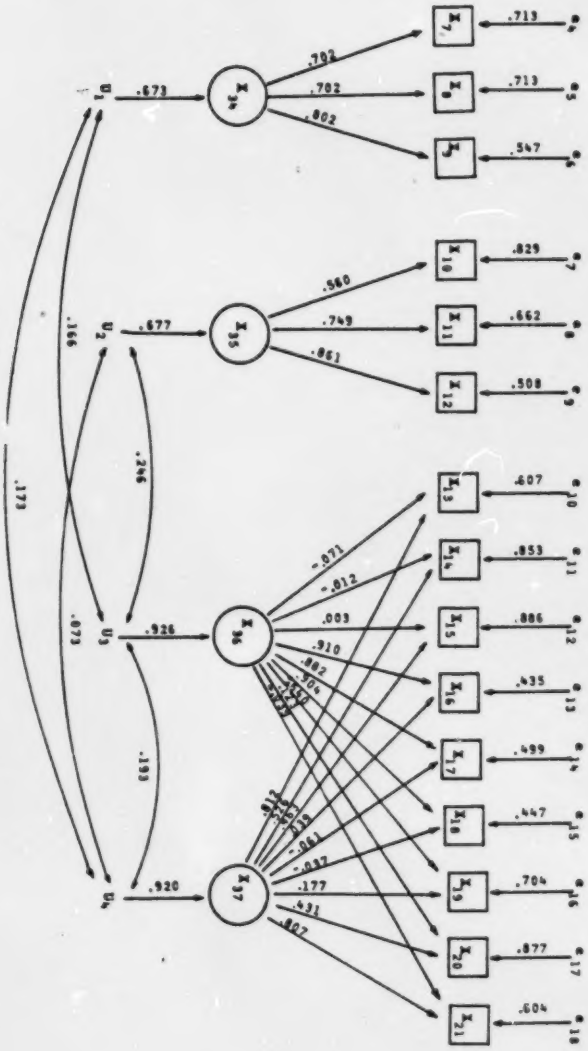
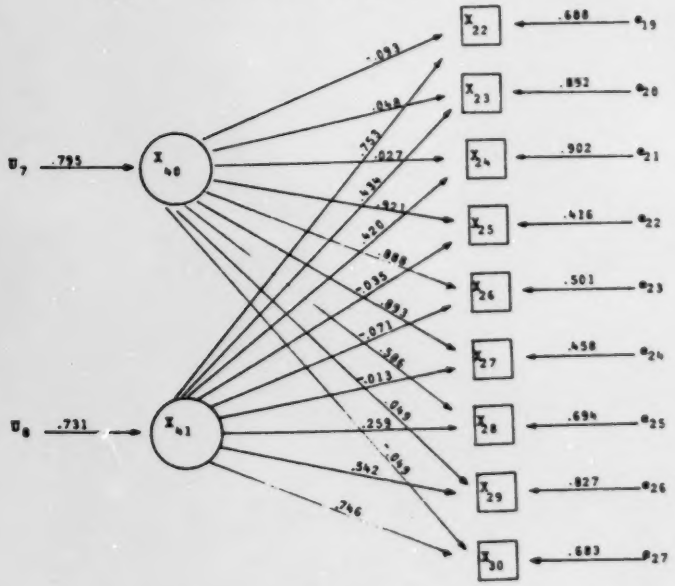


FIGURE 3. (continued)

Figure 3c. Outer Relations Among the Learning Outcomes



a. Latent variables are circled; indicator variables are in boxes. See Table 1 for a description of the variables.

TABLE 1. VARIABLES IN THE ESTIMATED BLOOM (1976) MODEL OF SCHOOL LEARNING

Variable Set	Manifest/Indicator Variables	Exogenous Source Variables		Endogenous Variables	
		Observed	Latent	Observed	Latent
Antecedent/ Background Variables	X_1 Father's occupation X_2 Father's education X_3 Mother's education	X_{32} Sex	X_{31} Socioeconomic Status		
	X_4 CAAT1 Verbal reasoning 9 X_5 CAAT2 Mathematical reasoning 9 X_6 CAAT3 Non-verbal reasoning 9		X_{33} IQ9		
Entry Behaviours	X_7 CEAT1 Reading Achivement 9 X_8 CEAT2 Mechanics of Expression 9 X_9 CEAT3 Effectiveness of expression 9				X_{34} Literacy 9
	X_{10} CMAT1 Arithmetic computation 9 X_{11} CMAT2 Facts, terms, and concepts 9 X_{12} CMAT3 Measurement 9				X_{35} Numeracy 9
	X_{13} Parent's suggestion re education & work 9 X_{14} Teacher's suggestion re education & work 9 X_{15} Friend's suggestion re education & work 9				X_{36} Motivation 9
	X_{16} Teacher's rating on reliability 9 X_{17} Teacher's rating on cooperation 9 X_{18} Teacher's rating on industry 9				
	X_{19} Teacher's rating--completing grade 13 9 X_{20} Student's intended occupation 9 X_{21} Student's own education & work plans 9				X_{37} Expectations 9
Learning Outcomes				X_{38} CATE English 10 X_{39} CATM Math 10	
	X_{22} Parent's suggestion re education & work 10 X_{23} Teacher's suggestion re education & work 10 X_{24} Friend's suggestion re education & work 10				X_{40} Motivation 10
	X_{25} Teacher's rating on reliability 10 X_{26} Teacher's rating on cooperation 10 X_{27} Teacher's rating on industry 10				
	X_{28} Teacher's rating--completing grade 13 (10) X_{29} Student's intended occupation 10 X_{30} Student's own education & work plans 10				X_{41} Expectations 10

TABLE 2. PATH COEFFICIENTS AND t-VALUES FOR OLS REGRESSION ANALYSIS

	Dependent Variables							
Independent Variables	X_{34} LIT9	X_{35} NUM9	X_{36} MOT9	X_{37} EXP9	X_{38} ENG10	X_{39} MATH10	X_{40} MOT10	X_{41} EXP10
	$\hat{\beta}$ t	$\hat{\beta}$ t	$\hat{\beta}$ t	$\hat{\beta}$ t	$\hat{\beta}$ t	$\hat{\beta}$ t	$\hat{\beta}$ t	$\hat{\beta}$ t
X_{31} SES	.064 2.76	-.016 -.68	.086 2.70	.175 5.53	.015 .68	.011 .44	-.034 -1.18	.146 5.48
X_{32} SEX	.143 6.36	-.077-3.38	.187 6.05	-.103-3.33	.179 8.05	-.079 -2.95	.066 2.23	-.003 -.11
X_{33} IQ	.721 31.22	.730 31.16	.318 9.97	.249 9.29	.237 6.50	.326 7.43	-.046 -.95	.136 3.07
X_{34} LIT9					.427 13.36	.075 1.94	.157 3.69	.056 1.44
X_{35} NUM9					.067 2.11	.230 6.01	.113 2.68	.069 1.78
X_{36} MOT9					.117 4.93	.172 6.02	.433 13.76	.095 3.31
X_{37} EXP9					.073 3.15	-.003 -.10	.030 .98	.426 15.02
R^2	.5531	.5403	.1482	.1583	.6182	.4458	.3247	.4343
Residual	.668	.678	.923	.917	.618	.744	.822	.752
Dmax	.042	.042	.042	.042	.715	.715	.715	.715

TABLE 3. PATH COEFFICIENTS AND t-VALUES FOR 2SLS REGRESSION ANALYSIS

	Dependent Variables							
Independent Variables	X_{34} LIT9	X_{35} NUM9	X_{36} MOT9	X_{37} EXP9	X_{38} ENG10	X_{39} MATH10	X_{40} MOT10	X_{41} EXP10
	$\hat{\beta}$ t	$\hat{\beta}$ t	$\hat{\beta}$ t	$\hat{\beta}$ t	$\hat{\beta}$ t	$\hat{\beta}$ t	$\hat{\beta}$ t	$\hat{\beta}$ t
X_{31} SES	.058 2.32	-.013 -.51	.086 2.70	.175 5.53	-.128 -1.80	.070 1.34	-.112 -2.72	.194 4.65
X_{32} SEX	.122 4.50	-.062 -2.33	.187 6.05	-.103 -3.33	.308 4.32	-.160 -3.68	.137 2.67	-.166 -2.27
X_{33} IQ	.902 7.23	.787 11.04	.318 9.97	.294 9.29	-.210 -1.22	.338 5.15	-.269 2.82	.287 1.96
X_{34} LIT9	—	-.080 -.86			.438 5.61	-.091 -1.12	.138 1.87	-.193 -2.07
X_{35} NUM9	-.246 -1.47	—			-.136 -1.37	.169 3.31	-.066 -.87	.072 .75
X_{36} MOT9					.221 1.44	-.130 -1.15	.299 5.59	-.323 -2.77
X_{37} EXP9					-.230 -1.87	.093 .99	-.179 -2.31	.386 8.75
X_{38} ENG10					—	.175 1.30	-.124 -.87	.269 1.63
X_{39} MATH10					.994 2.38	—	.629 2.32	-.580 -1.44
X_{40} MOT10					-.776 -1.70	.690 2.49	—	1.106 3.51
X_{41} EXP10					.733 2.58	-.277 -1.25	.483 2.84	—
R^2	.4925	.5225	.1482	.1582	-.3395	.2154	.1762	-.2338
Residual	.712	.691	.923	.917	1.157	.886	.908	1.111
Dmax	.562	.570	.042	.042	.739	.733	.745	.747

TABLE 4. THE ESTIMATED BLOOM 2SLS MODEL OF SCHOOL LEARNING[a]

Grade nine literacy:

(1) $X_{34_2} = -.307 - .256\,X_{35} + .049\,X_{31} + .210\,X_{32} + .849\,X_{33}$

$R^2 = .493$

Grade nine numeracy:

(2) $X_{35_2} = .278 - .077\,X_{34} - .010\,X_{31} - .103\,X_{32} + .712\,X_{33}$

$R^2 = .523$

Grade nine motivation:

(3) $X_{36_2} = .851 - .074\,X_{31} + .081\,X_{32} + .253\,X_{33}$

$R^2 = .148$

Grade nine expectations:

(4) $X_{37_2} = .315 + .125\,X_{31} - .148\,X_{32} + .232\,X_{33}$

$R^2 = .158$

Grade 10 Achievement in English Language:

(5) $X_{38} = 1.904 + 2.585\,X_{39} + 13.887\,X_{41} - 11.216\,X_{40} - 1.734\,X_{31}$
(21.576) (1.0848) (5.3879) (6.5974) (.9617)

$+ 8.416\,X_{32} - 3.133\,X_{33} + 6.934\,X_{34} - 2.231\,X_{35}$
(1.9502) (2.5744) (1.2367) (1.6305)

$+ 3.480\,X_{36} - 4.394\,X_{37}$
(2.4087) (2.3509)

$R^2 = -.340$

Grade 10 Achievement in Mathematics:

(6) $X_{39} = 16.203 + .067\,X_{38} - 2.020\,X_{41} + 3.832\,X_{40} + .363\,X_{31}$
(2.8191) (.0515) (1.6213) (1.5366) (.2710)

$- 1.680\,X_{32} + 1.939\,X_{33} - .557\,X_{34} + 1.071\,X_{35}$
(.4563) (.3764) (.4983) (.3237)

$- .785\,X_{36} + .687\,X_{37}$
(.6821) (.6945)

$R^2 = .215$

Grade 10 Motivation:

(7) $X_{40} = -1.949 - .009\,X_{38} + .113\,X_{39} + .633\,X_{41} - .105\,X_{31}$
(.9124) (.0099) (.0487) (.2228) (.0386)

$+ .259\,X_{32} - .277\,X_{33} + .151\,X_{34} - .076\,X_{35}$
(.0969) (.0984) (.0806) (.0869)

$+ .325\,X_{36} - .236\,X_{37}$
(.0582) (.1023) $R^2 = .176$

Grade 10 Expectations:

(8) $X_{42} = .980 + .014\,X_{38} - .079\,X_{39} + .843\,X_{40} + .138\,X_{31}$
(1.0512) (.0087) (.0554) (.2400) (.0297)

$- .240\,X_{32} + .225\,X_{33} - .161\,X_{34} + .062\,X_{35}$
(.1056) (.1149) (.0779) (.0832)

$- .268\,X_{36} - .268\,X_{37}$
(.0444) (.0968)

$R^2 = -.234$

a. The estimated betas are the regression coefficients. See Table 3 for path coefficients (i.e., standardized partial regression coefficients) and t-values. The numbers in parentheses for equations 5 through 8 are the standard errors.

TABLE 5. PATH COEFFICIENTS AND t-VALUES FOR SIMPLE NORMALIZATION REGRESSION ANALYSIS

	Dependent Variables							
Independent Variables	X_{34} LIT9	X_{35} NUM9	X_{36} MOT9	X_{37} EXP9	X_{38} ENG10	X_{39} MATH10	X_{40} MOT10	X_{41} EXP10
	$\hat{\beta}$ t	$\hat{\beta}$ t	$\hat{\beta}$ t	$\hat{\beta}$ t	$\hat{\beta}$ t	$\hat{\beta}$ t	$\hat{\beta}$ t	$\hat{\beta}$ t
X_{31} SES	.066 2.91	-.012 -.52	.086 2.75	.173 5.58	.023 1.21	.016 .72	-.026 -1.03	.138 6.04
X_{32} SEX	.139 6.32	-.076 -3.41	.181 6.04	-.101 -3.35	.165 8.80	-.071 -3.14	.070 2.81	-.010 -.46
X_{33} IQ	.706 31.27	.715 31.20	.311 9.99	.289 9.33	.226 9.77	.273 9.77	-.004 -.14	.121 4.30
X_{34} LIT9					.367 16.40	.103 3.82	.132 4.42	.074 2.72
X_{35} NUM9					.093 4.20	.224 8.34	.101 3.42	.076 2.80
X_{36} MOT9					.117 5.96	.154 6.49	.378 14.44	.095 3.96
X_{37} EXP9					.076 3.87	.011 .45	.037 1.42	.372 15.64
R^2	.5320	.5194	.1426	.1532	.5480	.3983	.2722	.3782
Residual	.684	.693	.926	.920	.627	.776	.853	.788
k_N	.020	.020	.020	.020	.138	.138	.138	.138
Dmax	.012	.012	.012	.012	.156	.156	.156	.156

TABLE 6. PATH COEFFICIENTS AND t-VALUES FOR 2SNR ANALYSIS

Independent Variables	Dependent Variables							
	X_{34} LIT9	X_{35} NUM9	X_{36} MOT9	X_{37} EXP9	X_{38} ENG10	X_{39} MATH10	X_{40} MOT10	X_{41} EXP10
	$\hat{\beta}$ t	$\hat{\beta}$ t	$\hat{\beta}$ t	$\hat{\beta}$ t	$\hat{\beta}$ t	$\hat{\beta}$ t	$\hat{\beta}$ t	$\hat{\beta}$ t
X_{31} SES	.072 3.44	-.016 -.75	.086 2.75	.173 5.58	.002 .114	.004 .20	-.053 -2.22	.147 6.71
X_{32} SEX	.145 7.14	-.100 -4.86	.183 6.04	-.101 -3.35	.165 8.82	-.107 -4.80	.066 2.79	-.042 -1.91
X_{33} IQ	.422 21.43	.484 19.96	.311 9.99	.289 9.33	.142 6.62	.229 8.49	-.076 -2.69	.060 2.37
X_{34} LIT9	—	.254 8.17			.307 13.56	.037 1.41	.071 2.51	-.011 -.42
X_{35} NUM9	.343 12.98	—			.019 .83	.191 7.27	.040 1.43	.025 .97
X_{36} MOT9					.009 .47	.054 2.40	.317 12.56	-.020 -.84
X_{37} EXP9					-.003 -.15	-.021 -.94	-.039 -1.60	.346 15.07
X_{38} ENG10					—	.115 3.73	.082 2.57	.127 4.15
X_{39} MATH10					.206 8.75	—	.172 5.69	.080 3.05
X_{40} MOT10					.161 5.46	.191 5.29	—	.225 6.08
X_{41} EXP10					.181 6.25	.063 1.92	.160 4.30	—
R^2	.5473	.5406	.1426	.1532	.6107	.4520	.3681	.4662
Residual	.673	.678	.926	.920	.624	.740	.795	.731
k_N	.115	.118	.020	.020	.154	.160	.163	.160
Dmax	.118	.109	.012	.012	.130	.109	.119	.126

Table A

Path Coefficients and their t-values (in parentheses) for Determinants of Grade 10 English Language Achievement: OLS Regression Model

Dependent Variable: Grade 10 English Language Achievement

	X_{31} SES	X_{32} SEX	X_{33} IQ	X_{34} LIT9	X_{35} NUM9	X_{36} MOT9	X_{37} EXP9
1. OLS	0.015 (0.68)	0.179 (8.05)	0.237 (6.50)	0.428 (13.36)	0.067 (2.11)	0.117 (4.93)	0.073 (3.15)
2. Normalization[a] (Lee 1979)	0.023 (1.12)	0.165 (8.80)	0.226 (9.77)	0.367 (16.40)	0.093 (4.20)	0.117 (5.96)	0.076 (3.87)
3. Kasarda & Shih[b] (1977)	0.017 (0.82)	0.176 (8.28)	0.234 (7.35)	0.410 (14.19)	0.075 (2.63)	0.117 (5.19)	0.074 (3.34)
4. Hocking, Speed & Lynn[c] (1976)	0.015 (0.69)	0.178 (8.07)	0.237 (6.57)	0.426 (13.42)	0.068 (2.15)	0.117 (4.95)	0.073 (3.17)
5. Hoerl, Kennard & Baldwin[d] (1975)	0.016 (0.72)	0.178 (8.13)	0.236 (6.77)	0.422 (13.62)	0.070 (2.28)	0.117 (5.01)	0.074 (3.21)
6. Lawless & Wang[e] (1976)	0.015 (0.70)	0.178 (8.09)	0.236 (6.63)	0.425 (13.48)	0.068 (2.19)	0.117 (4.97)	0.074 (3.18)
7. Hoerl & Kennard's Iterative[f] (1976)	0.016 (0.72)	0.178 (8.13)	0.236 (6.77)	0.422 (13.62)	0.070 (2.28)	0.117 (5.01)	0.074 (3.21)
8. Vinod[g] (1976)	0.034 (2.85)	0.115 (9.77)	0.185 (18.13)	0.254 (23.56)	0.115 (10.61)	0.108 (9.29)	0.073 (6.20)

Computing formulae:

a. Normalization: $\frac{1}{p}\Sigma \frac{\lambda_i}{(\lambda_i + k)^2} = 1$

b. Kasarda & Shih: $MSE = \hat{\sigma}^2 \Sigma \frac{\lambda_i}{(\lambda_i + k)} + k\Sigma \frac{\hat{\alpha}_i^2}{(\lambda_i + k)} = \text{minimum}$

c. Hocking et al.: $k = \hat{\sigma}^2 \frac{\Sigma \lambda_i^2 \hat{\alpha}_i^2}{\Sigma \lambda_i^2 \alpha_i^4}$

d. Lawless & Wang: $k = \frac{p\hat{\sigma}^2}{\Sigma \lambda_i \hat{\alpha}_i^2}$

e. Hoerl, Kennard, & Baldwin: $k = \frac{p\hat{\sigma}^2}{\Sigma \hat{\alpha}_i^2}$

f. Hoerl & Kennard's iterative: $k_0 = \frac{p\hat{\sigma}^2}{\Sigma \hat{\alpha}_i^2}$;

$k_t = \frac{p\hat{\sigma}^2}{\Sigma \hat{\alpha}_{it}^{*2}}$, $\hat{\alpha}_{it}^{*} = \frac{\lambda_i \hat{\alpha}_i}{(\lambda_i + k_t)}$, with tolerance level $\frac{k_{t+1} - k_t}{k_t} \leq \delta$

where $\delta = 20T^{-1.38}$; $T = tr\,(X'X)^{-1}/p$

g. Vinod: $ISRM = \Sigma\{(p\lambda_i/(\lambda_i + k)^2 S) - 1\}^2 = \min$; where $S = \Sigma \lambda_i/(\lambda_i + k)^2$

Table B. The Characteristics and Performance Indices of OLS and Ridge Regression Procedures for the OLS Grade 10 English Language Achievement Equation

($p = 7$, $N = 896$, $D_c = 0.830$)

Dependent Variable: Grade 10 English Language Achievement

		k[a]	Vmax[b]	Dmax[c]	E.S.Ratio[d]	RVAR[e]	ϕ_0[f]	α[g]	R^2[h]
1.	OLS	0	3.09	.715	1.00	0	0	1.00	.618
2.	Normalization	.138	1.25	.156	.405	44.59%	.86%	.363	.548
3.	Kasarda & Shih	.033	2.37	.598	.764	15.84%	.06%	.999	.599
4.	Hocking et al.	.002	3.02	.707	.978	1.39%	.00%	1.00	.617
5.	HKB	.010	2.83	.682	.915	5.47%	.01%	1.00	.612
6.	Lawless & Wang	.005	2.96	.700	.958	2.68%	.00%	1.00	.615
7.	H&K's Iterative	.010	2.83	.682	.915	5.47%	.01%	1.00	.612
8.	Vinod	.741	3.27	-0.377	.065	83.59%	10.83%	-0.00	.364

a. The bias parameter k (for details see footnotes to Table A)

b. The maximum VIF (see footnote #6)

c. The D-measure of multicollinearity has a range of 0 to 1 (see footnote #6)

d. The empirical sensitivity ratio $= \frac{SSF^*}{SSF} = \frac{\Sigma(\hat{\beta}_i^* - \underline{\beta}_i^*)^2}{\Sigma(\hat{\beta}_i - \underline{\beta}_i)^2}$; where $\underline{\beta}_i$ and $\underline{\beta}_i^*$ represent the ith perturbed OLS regression coefficient and that of the ridge regression coefficient. The amount of perturbation $|\Delta r| = 0.01$ was introduced to the largest r coefficient in the r matrix.

e. Relative reduction in variance $= 1 - \frac{tr(VIF)}{tr(X'X)^{-1}}$

f. Inflation in the residual sum of squares.

g. Acceptance level of ridge matrix; i.e., $F_{p,\, n-p,\, (1-\alpha)} = \frac{(\hat{\beta} - \hat{\beta}^*)'X'X(\hat{\beta} - \hat{\beta}^*)}{p\alpha^2}$ (see McCabe, 1978).

h. The multiple R square.